



中国开源软件推进联盟
China Open Source Software Promotion Union



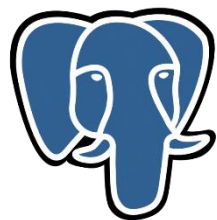
第十二届开源中国开源世界高峰论坛 中国开源联盟PostgreSQL分会成立发布会 暨 2017年PostgreSQL技术论坛

加拿大渥太华 PGCon 2017 归来

周宝峰

瀚高北美研发中心

2017年6月21日



PostgreSQL



www.highgo.com

■ 关于 周宝峰 (Grant Zhou)

瀚高在美国西雅图地区的北美研发中心负责人

美国官方PostgreSQL协会Professional Member

✧ 自2000年参加工作以来，先后在朗讯科技、大富科技、乾程科技、Corinex、Fortinet工作过十余载，技术出身，也是开源技术的爱好者。

✧ 现在专注于PostgreSQL的企业级应用及云计算、大数据业务

✧ 作为一名非资深PostgreSQL玩家，希望以后能跟PostgreSQL社区一起多交流学习！



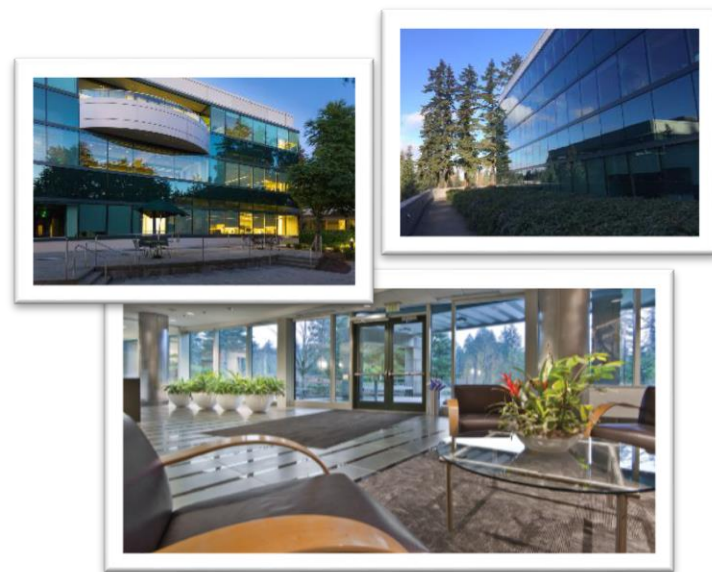
✉ grantzhou@highgo.com in grantzhou

▪ ▪ ▪ www.highgo.com

■ 瀚高美国研发中心

瀚高立志于要做社区最好的PostgreSQL发行版

- 引进美国的先进技术和理念，与国内PG社区一起实现这个目标
- 2016年成立瀚高北美研发中心（@ Bellevue Technology Center），加强国际合作与沟通
- 借助国际社区力量，协作研发适合国内本土的PG功能，增强国内PG实力



▪ ▪ ▪ www.highgo.com

关于PGCON 2017

PGCon是PostgreSQL用户和开发者的年度大会，每年都在加拿大的渥太华大学内举行。

今年PGCon已经是第十二次举行，这次大会参加人员两百多人，主要参与人员是PostgreSQL社区的开发人员及使用用户

链接 <http://www.pgcon.org/2017/>

会议演示PPT瀚高已经打包，大家可以到这个地址下载：**会后更新**



PGCon 2017

A conference for PostgreSQL
Users and Developers

关于PGCON 2017的主要参与企业

- 今年的赞助商包括



里面的大部分PG相关企业是每次PG会议上的熟悉面孔

- PGCON以技术演讲为主，现场不组织企业宣传活动
相对而言美国的PG CONF、PG Vision、PG Open等会议，商业氛围偏弱，但技术氛围更浓厚一些
- 国内企业华为在2015年参与过，其他企业很少出现在PGCON上，瀚高在今年的大会上分享了我们的XLogMiner的开源项目

www.highgo.com



■ 瀚高 和 PG 社区

瀚高一直积极的参与到PG社区的活动中，自2016年北美研发中心建立之后，公司更是加大了跟国际社区的交流，参加了2016年的美国洛杉矶PGVision 2016, 2017年美国泽西市的美国PG用户大会（PGConf US），加拿大渥太华的全球用户与开发者大会（PGCon），以及计划中的今年美国洛杉矶PG Open 2017，PGVision 2017，美国四大本地PGConf之西雅图PGConf等。

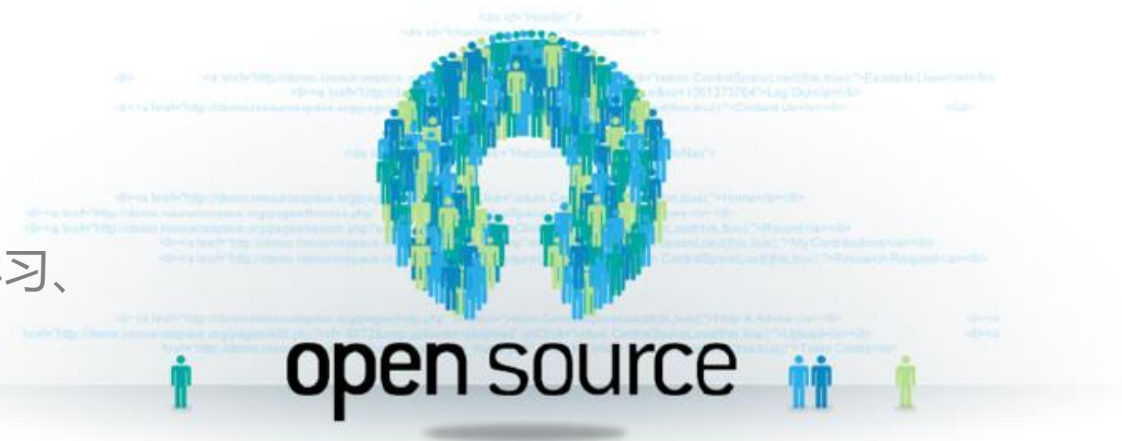
为什么瀚高要积极的参与到北美地区的这些会议中去？

与PG开发者/用户/相关企业能更有效的进行交流、沟通、学习、拓宽视野并期待以后在社区中做出更多的贡献！

因开源而收获，亦要回馈开源，反哺社区！

拥抱 回馈 融合 开放

www.highgo.com



PGCON 2017 全球用户与开发者大会 小结

- 参与企业情况

美国企业 No.1，其他如俄罗斯、瑞典、法国、日本等

日本在PostgreSQL社区的影响力不断加强

- 商业活动相对较少，参与的社区开发者较多，讨论的议题也较为深入

- 除了传统的针对PG使用、性能调优、高可用、监控管理、安全等方面的演讲及培训外，此次会议也开带来了有关云上（Cloud）高可用PosrgreSQL、AI和机器学习方面的议题，同时也深入讲述了关于今年即将发布的PG10的新特性

■ PostgreSQL 技术发展趋势

- PG 发展趋势
 - 安全功能、性能、多环境部署、NoSQL支持目前仍然是PG发展的重点趋势
 - 多环境支持，特别是云环境，目前各个PG商业公司都已发行了其容器版本。同时，基本所有的云提供商也都已经提供了对PostgreSQL的支持

PG未来发展动向和计划 – PG 10 版本

详细PG10特性清单请参考PG官网：

<https://www.postgresql.org/docs/devel/static/release-10.html>

PostgreSQL继续向着将在今9月份发布的第10个主版本迈进，PG 10 Beta 1 在2017年5月18号刚刚发布。
PG10值得关注的几个新特性：

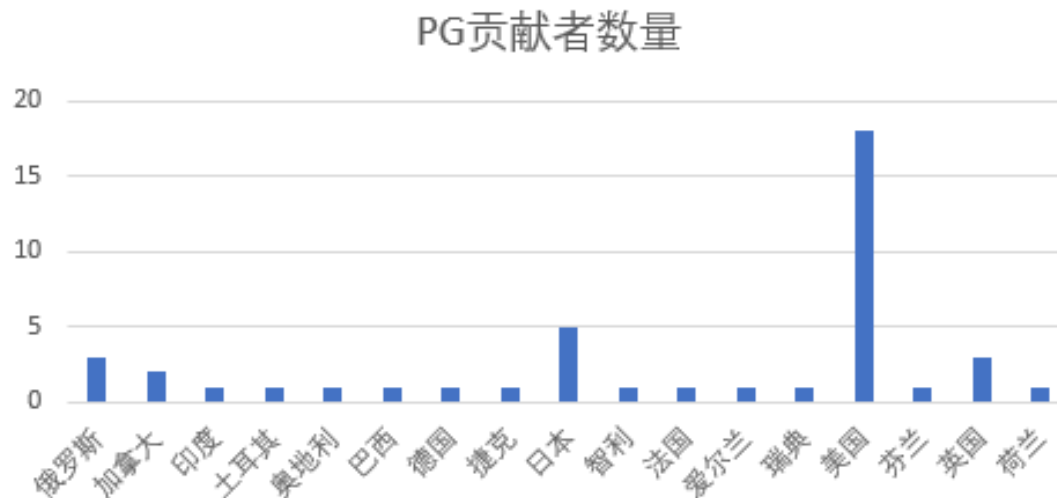
- 逻辑复制：表级粒度（table-level granularity）、允许多集群对单集群复制、单表对多集群复制、跨PG主版本间复制、在订阅者（subscribers）端创建本地对象（如表索引）等。
- 改进并行查询 - 2到4倍的速度提升
- 更强大的密码哈希算法 - 使用了SCRAM-SHA-256
- 表分区 - 表继承的简化版本

其他比如自治事务、多主集群、使用分片提高扩展性和可用性、查询JIT编译、可插拔的存储引擎（columnar、内存内存存储等）、页级数据压缩、改进临时表性能等也是值得大家期待的几个特性。

▪ ▪ ▪ www.highgo.com

PostgreSQL 在世界各地的推广和应用

- 美国继续引领着PG的开发，在全球PG相关的商业市场上继续保持着一枝独秀。随着云服务的继续普及，PG在云上的部署更加普遍，一个很明显的现象就是Amazon携带其AWS/RDS/EC2等相关服务的应用介绍非常频繁的出现在各个PG会议上
- 日本目前不管是市场方面还是开发贡献方面，一直走在整个亚洲的前面，比如NTT公司的OSS部门以及日本的EDB分公司等，在PG10的开发中承担了很多新功能的开发工作，比如FDW部分功能、同步复制等
- 欧洲方面，以英国为首，其他分散在德国、法国、奥地利等地区
- 在俄罗斯还是以Postgres Professional公司为主导

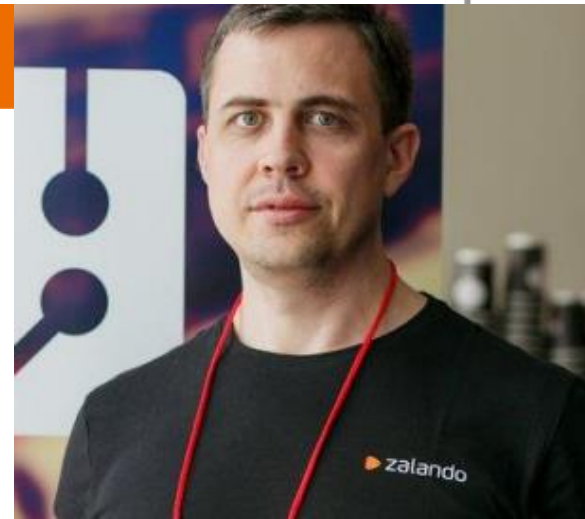


收获与成果几个有特色的演讲、话题

如何使用Docker, Spilo 和 Patroni 建立云上高可用的PostgreSQL

演讲者：Patroni项目的主要贡献者，德国的Zalando SE公司的Alexander Kukushkin

资料地址：http://www.pgcon.org/2017/schedule/attachments/448_PGCon-2017.pdf



简介：如何使用Patroni管理运行在云端的数以百计的PostgreSQL数据库，包括采用分布式配置仓库（Etcd，Zookeeper或者Consul）来保持当前集群的状态，以及稳定的检查主端失败等。同时，也讲了如何将运行在AWS的部署迁移到运行在AWS之上的Kubernetes集群。

Patroni是一个模板，它使用Python为你提供一个自己订制的高可用的解决方案，配置信息存储在像ZooKeeper, etcd或者Consul中。它可以帮助你在数据中心或者任何地方快速部署高可用的PostgreSQL方案

<https://patroni.readthedocs.io/en/latest/>



收获与成果几个有特色的演讲、话题

如何在安全要求最高的环境中使用PostgreSQL数据库

演讲者： Masanori Oyama □ □ 真実

来自日本NTT开源中心数据库系统开发和支持团队

https://www.pgcon.org/2017/schedule/attachments/451_PostgreSQL%20Security.%20How%20Do%20We%20Think_pub.pdf

简介：讲了信用卡行业数据安全标准（PCI DSS）的要求，并讲述了如何在安全要求最高的环境中使用PostgreSQL数据库，以及相关的审计方面要求。

价值：对如何将PG应用到关键性行业业务系统中去提供了很好的参考，此次演讲详细讲述了信用卡数据安全标准的具体安全要求、如何使用PG现有功能进行配置支持、以及对未来PG需要开发的功能以更好的支持安全标准作出了建议。

■ 收获与成果 瀚高XLogMiner

XLogMiner是一个wal日志解析工具。它能根据当前的wal日志和数据库的数据字典，还原出产生这些wal日志的SQL语句，并针对DML语句生成undo语句。

我们这次在PGCON 2017上也将其开源，目前是基于PostgreSQL 9.5开发的。

下面地址

<https://github.com/HighgoSoftware/XLogMiner>

中文使用说明：

<https://github.com/HighgoSoftware/XLogMiner/blob/master/README.CN.md>

▪ ▪ ▪ www.highgo.com

XLOGMiner 基本原理



WAL logs



Current database

XLogMiner

xid	1952
virtualxid	1
timestampz	2017-05-23 14:48:51.895002+08
record_database	highgo
record_user	lchch
record_tablespace	pg_default
record_schema	public
op_type	UPDATE
op_text	UPDATE "public"."xlogminer_test" SET "att_name" = 'BBB_CHANGE' WHERE "att_id"=2 AND "att_name"='BBBBBBB'
op_undo	UPDATE "public"."xlogminer_test" SET "att_name" = 'BBBBBBB' WHERE "att_id"=2 AND "att_name"='BBB_CHANGE' AND ctid = '(0,4)';

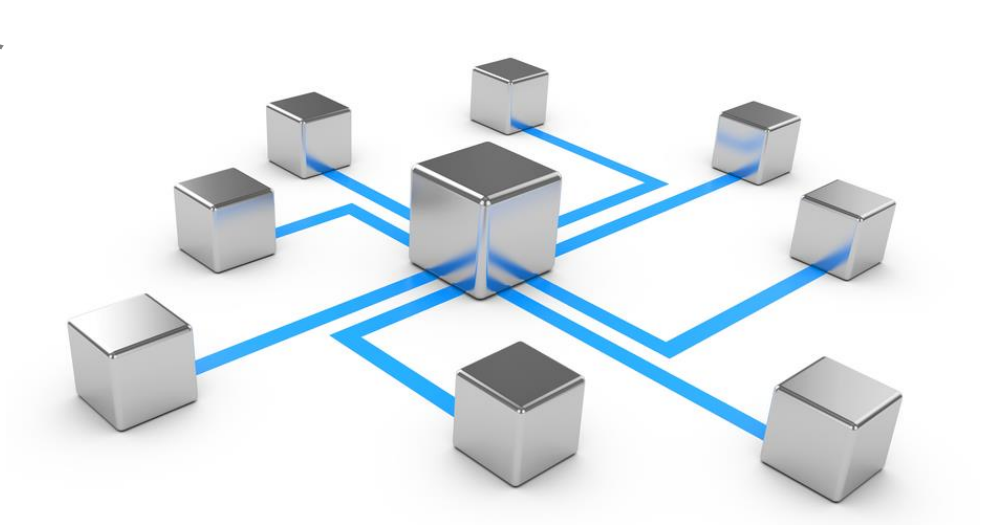
SQL

+

UNDO SQL

■ 共享存储

更多的企业从开源中获益，瀚高也不例外，我们基于PG开发了企业级数据库（HGDB）并应用在各个行业的解决方案中，同时我们也做了很多相关的工具，除了刚才讲的开源XLogMiner之外，瀚高根据行业关键业务需求研发了共享存储解决方案，希望能给大家带来一些新的思路



■ 共享存储系统的场景分析

集群系统一般运用在银行、证券等高关键性系统中，此类系统的要求

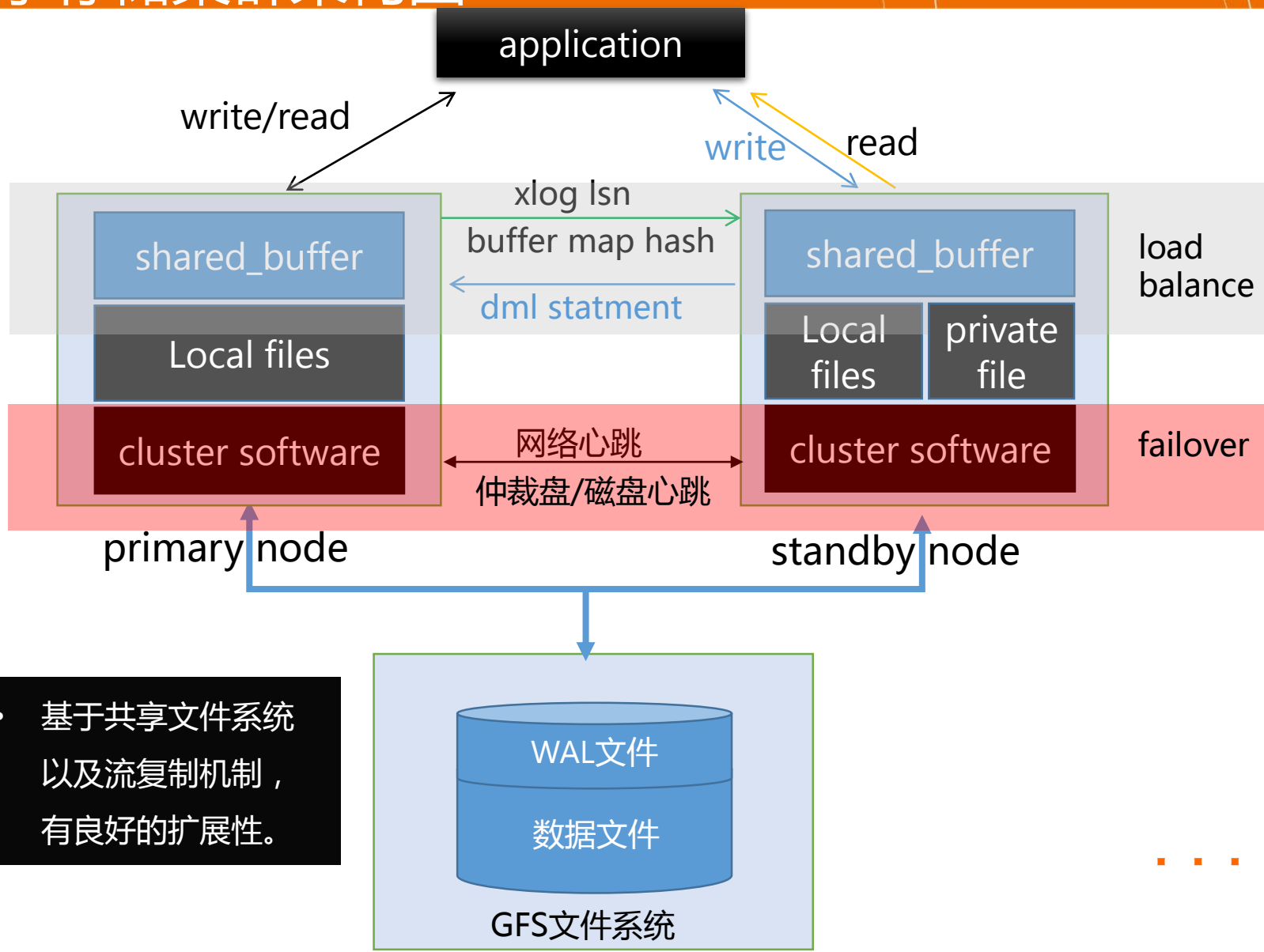
- 高稳定性
- 高性能
- 高并发
- 数据强一致性

而PG目前的高可用方案都难以完全满足以上需求，特别是如何解决数据一致性问题

下面，请 瀚高数据库内核研发经理 - 王硕 讲述一下具体技术细节

HGDB共享存储集群架构图

- 数据变更写入共享文件系统
- 利用xlog lsn读取本地wal文件，在内存中执行 recovery，完成脏数据同步

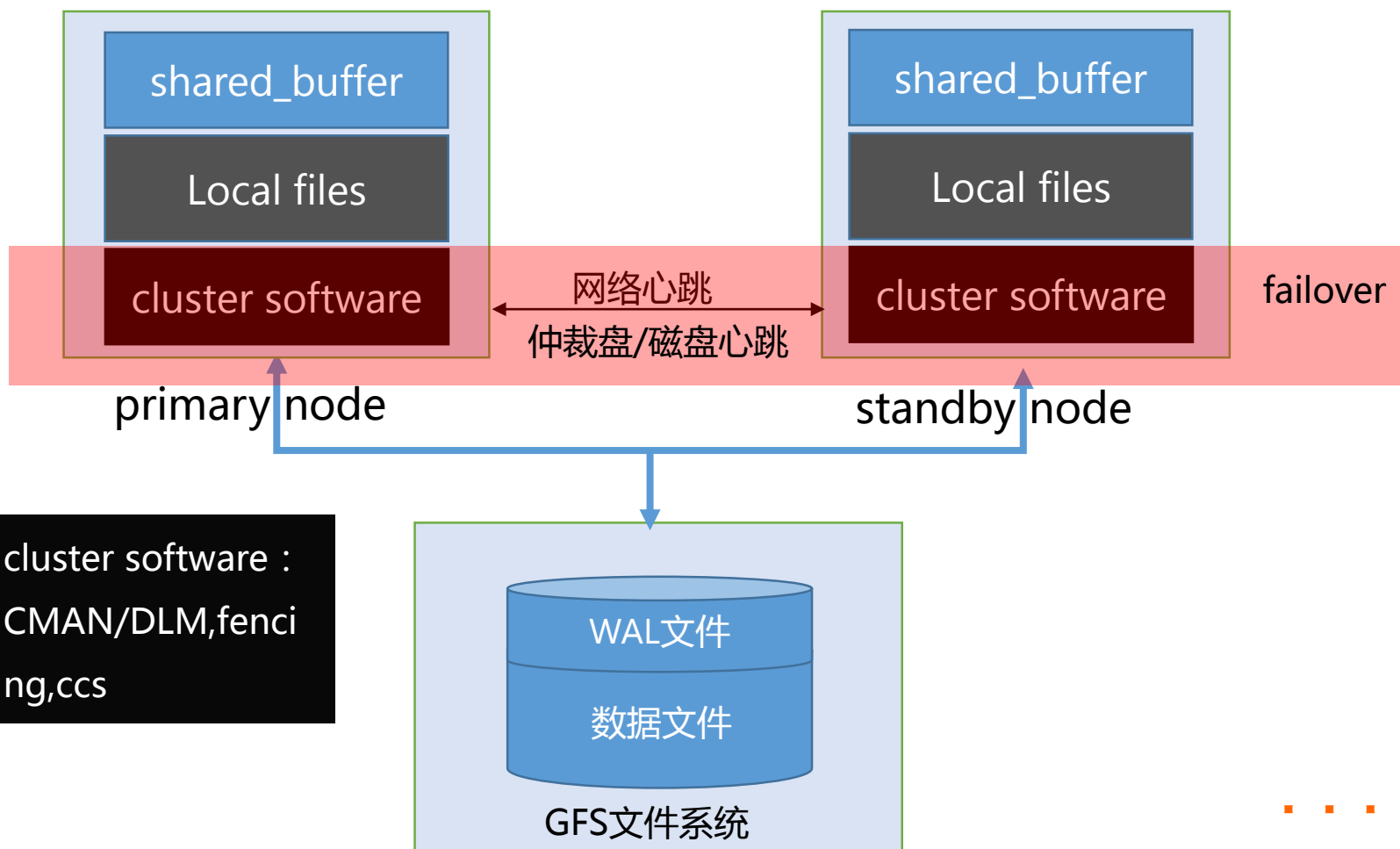


- 主节点支持读写操作，本地完成
- 从节点支持读写操作，读本地完成，写操作转发到主节点

- 基于共享文件系统以及流复制机制，有良好的扩展性。

■ HGDB共享存储集群文件系统

- 为保证I/O独立性，使用GFS集群文件系统。

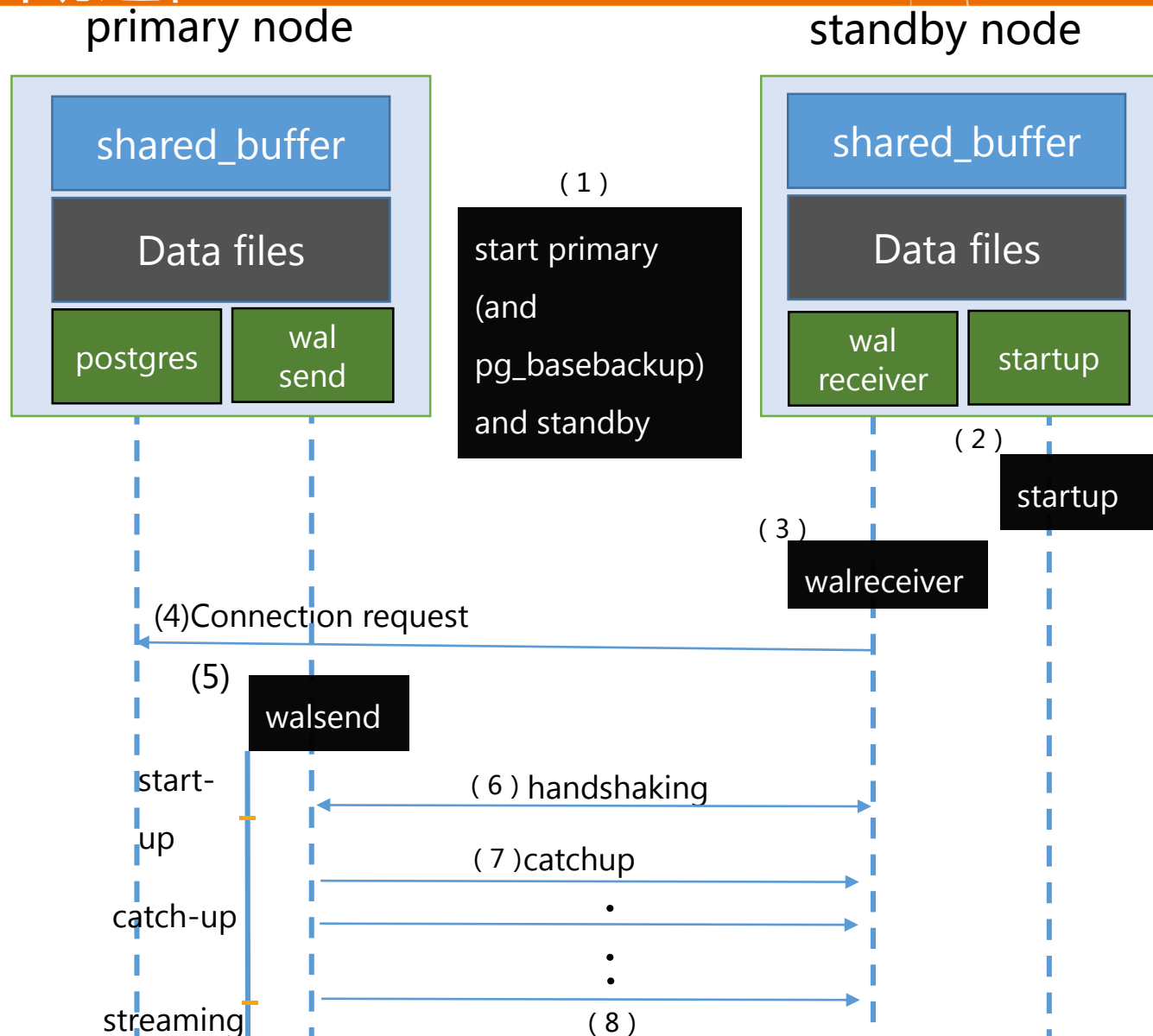


- cluster software : CMAN/DLM,fencing,ccs

- 共享存储集群支持主备两种模式，所以必然存在不同的文件服务不同的节点，因此在共享存储之外还存在各节点的本地文件。

PG流复制启动过程

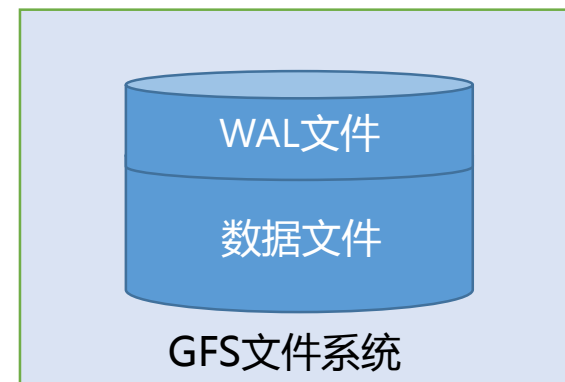
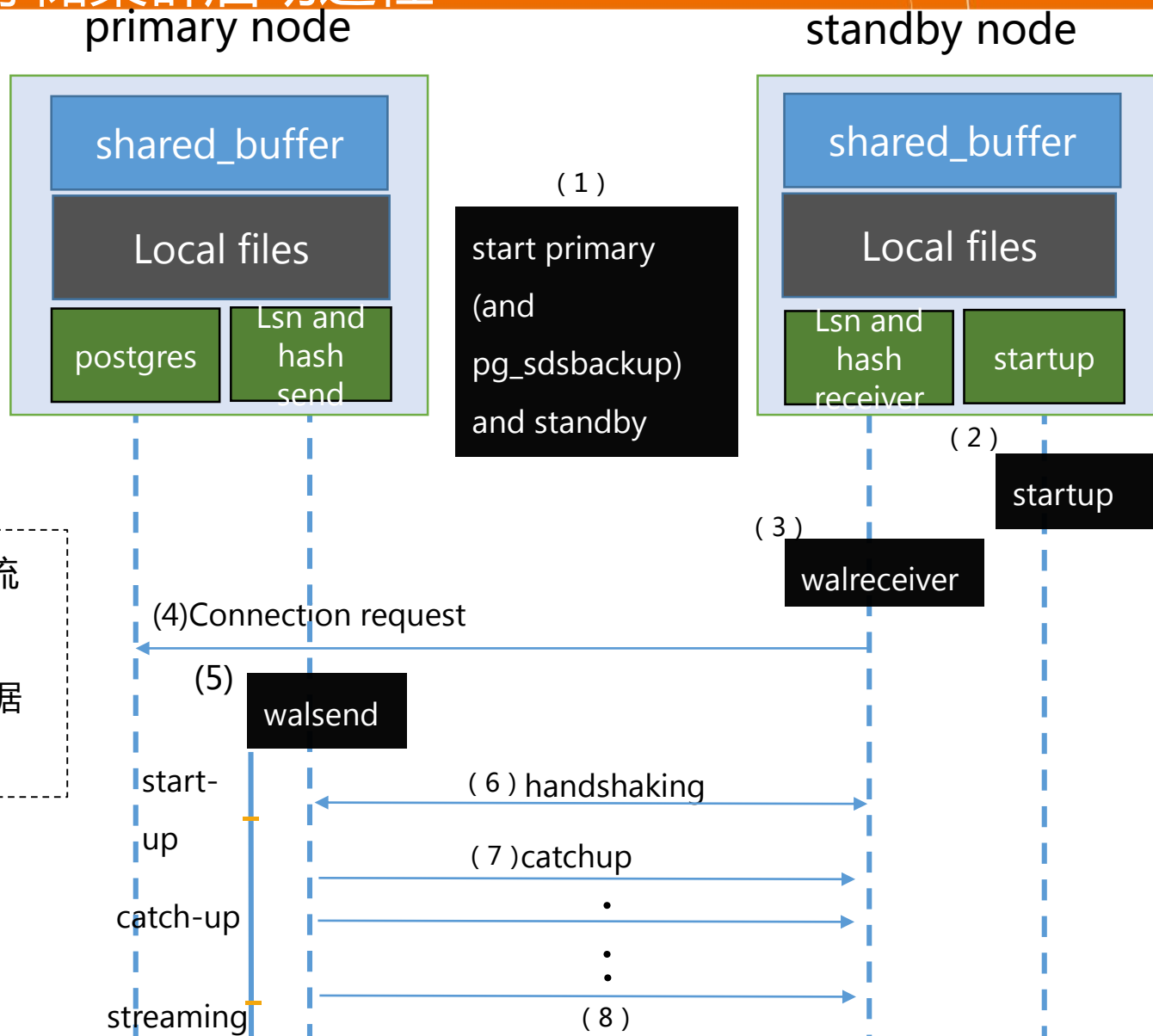
原有流复制原理：
复制wal到备端，然后
重演操作，完成更新。



HGDB共享存储集群启动过程

pg_sdsbackup :
备端文件准备，不完
全复制原有文件，按
照需求复制或重写。

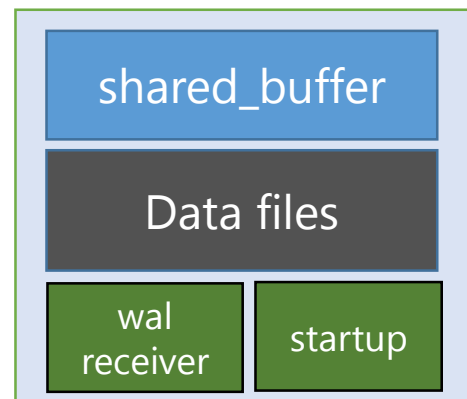
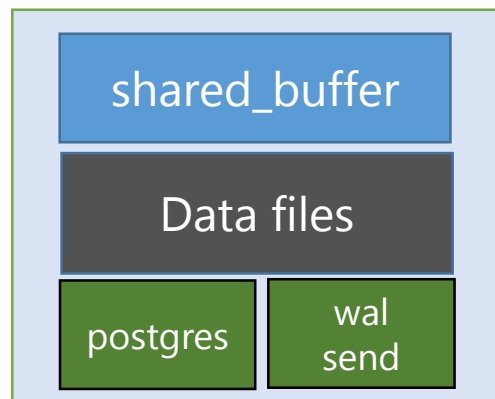
SDS流复制与一般流
复制区别：
追平过程仅为脏数据
追平。



PG同步流复制过程

primary node

standby node



(1) Write and flush wal

(2) sends the WAL data written into the WAL segment

(3) wait

(7)

ACK

(4)

ACK

(5)

ACK

(6)

ACK

heartbeat

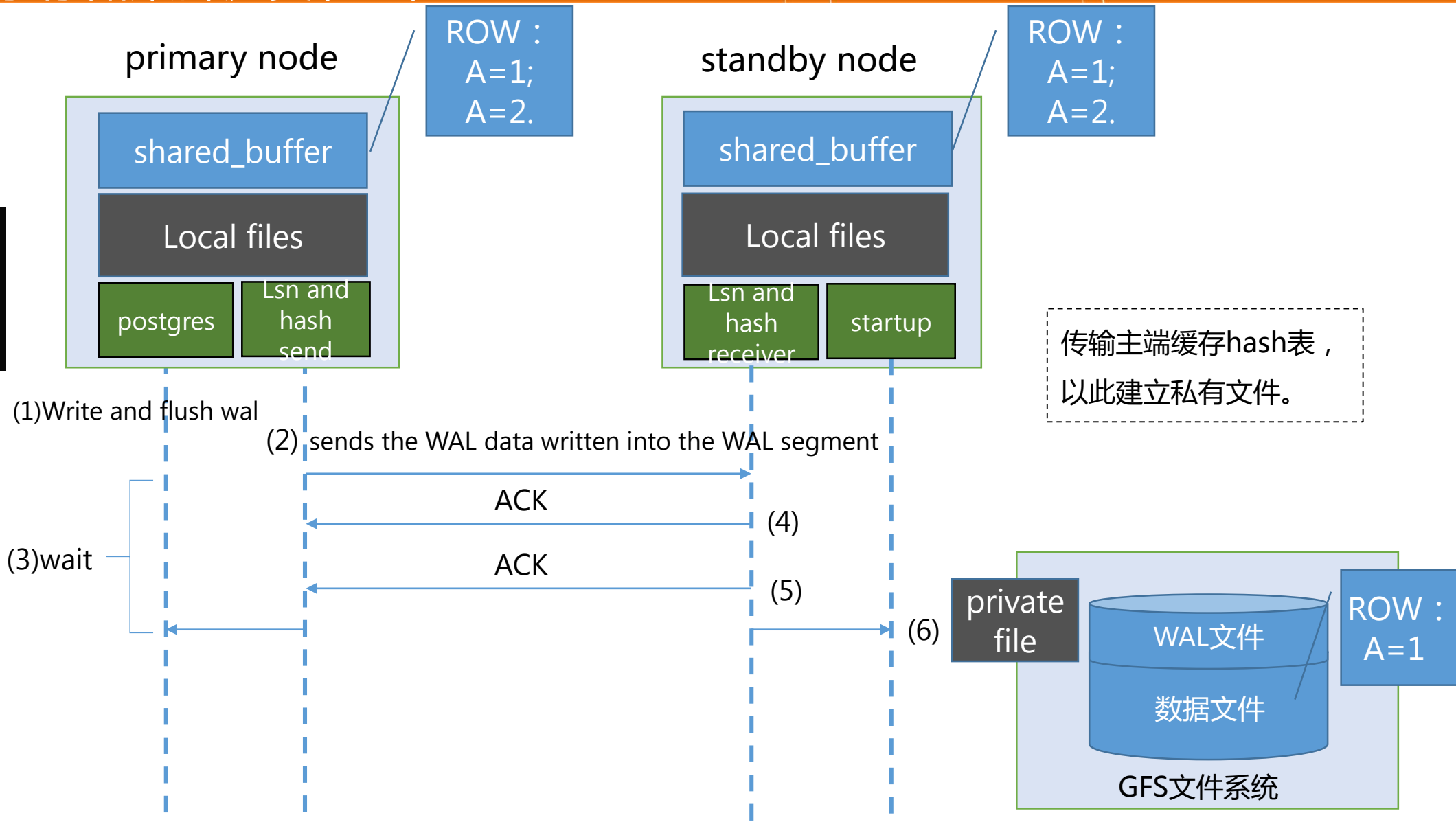
ACK信息：

已被写入的LSN位置信息。
已被刷入磁盘的LSN位置
信息。
已被重演的LSN位置信息。
时间。

HGDB共享存储集群流复制过程

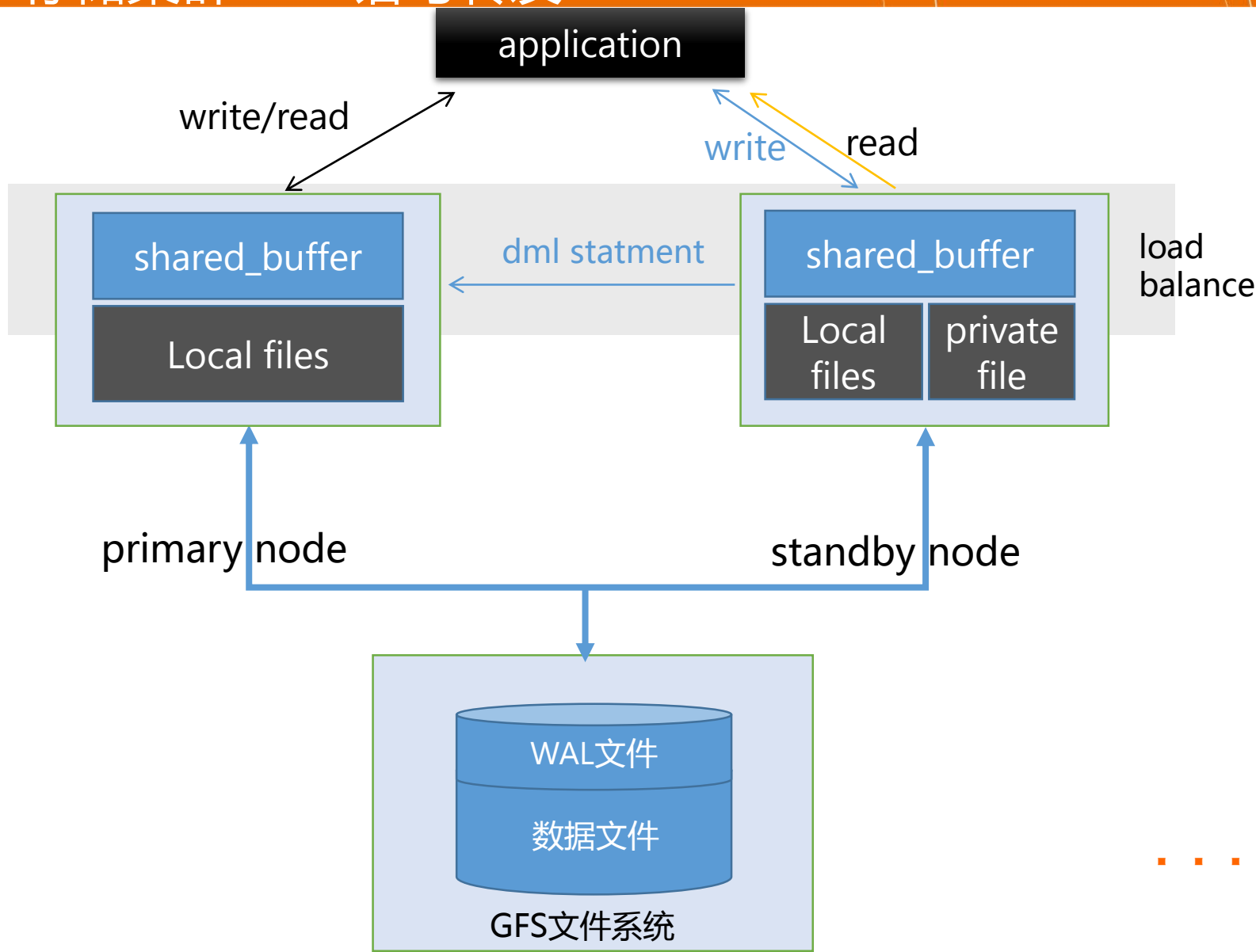
SDS流复制机制：
只发送LSN，备端查找wal日志后重演。

备端提供读服务，其次不可写，可能会造成dirty page被置换掉，因此创建私有文件保存此数据



HGDB共享存储集群DML语句转发

- 备节点识别SQL语句后，判定为写操作即转发到主端完成执行，而后主端发送确认信息至备端。



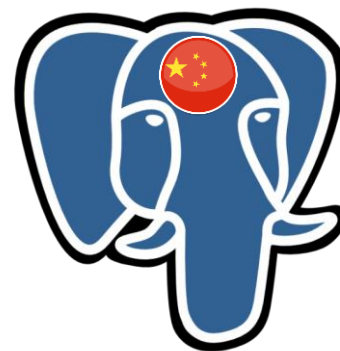
■ 结束语

每次PG会议，组织方都会邀请对社区作出相对较大贡献的开发人员或用户参加，目前我们国内的社区成员参加的次数还是很有限，希望我们国内社区成员更多的出现在邀请行列之内。

对于PG相关企业，希望能借助PG分会的影响，加大沟通与合作，抱团发展，一起为PG事业的建设提供帮助。

走出去 引进来 加强交流 合作共赢

PG开源社区需要来自**中国**的声音



PostgreSQL · · · www.highgo.com

CONTACT



2018 156th Ave NE, Bellevue, WA 98007, USA

www.highgo.com

